

Reliable Fidelity and Diversity Metrics for Generative Models

Year: 2020 | Citations: 488 | Authors: Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, Jaejun Yoo

Abstract

Devising indicative evaluation metrics for the image generation task remains an open problem. The most widely used metric for measuring the similarity between real and generated images has been the Frechet Inception Distance (FID) score. Because it does not differentiate the fidelity and diversity aspects of the generated images, recent papers have introduced variants of precision and recall metrics to diagnose those properties separately. In this paper, we show that even the latest version of the precision and recall metrics are not reliable yet. For example, they fail to detect the match between two identical distributions, they are not robust against outliers, and the evaluation hyperparameters are selected arbitrarily. We propose density and coverage metrics that solve the above issues. We analytically and experimentally show that density and coverage provide more interpretable and reliable signals for practitioners than the existing metrics. Code: [this https URL](https://github.com/mfnaeem/reliable_fidelity_and_diversity).